Credit Scoring of Bank Depositor with Clustering Techniques for Supply Chain Finance

Abdollah Nazari[1], Mohammadreza Mehregan [2], Reza Tehrani [3]

[1] *Production and Operations Management, Alborz College, University of Tehran*

[1]nazari_iee@yahoo.com

[2,3]*Faculty of Management, University of Tehran*

***Abstract*- One of the natural consequences of lending practices by banks and credit institutions has been the creation of deferred and doubtful loans- a phenomenon that has become a major concern for these institutions and has had a negative impact on their revenue and expenditure. From an internal perspective, operating costs, work efficiency, profitability, customer service, branch rank, employee wages and salaries, and other qualitative indicators are significantly affected. From an external perspective, these loans lead to slow cash flow, lack of timely and optimal allocation of resources to manufacturing networks and industries, low employment rates, and eventually economic recession. The purpose of this research is to cluster bank customers and determine the behavioral pattern of each cluster for supply chain finance using K-Means, FCM, and SUB Cluster models in Clementine 18.0, MATLAB 2016, and Excel software. 35 models were compared with a variety of parameters. After removing nonessential variables, the models were rerun and the outputs for each customer cluster were provided. The results showed that creditworthiness, education, job, collateral value, collateral type, loan term, and age respectively had the greatest impact. Finally, the K-Means model was found to be the most appropriate clustering technique.**

Keywords: *credit scoring, clustering Techniques, K-Means, FCM, supply chain finance, bank depositor.*

1. Introduction

Banks and financial institutions are intermediary institutions which receive money from customers having extra money and allocate it to the depositors looking for money for supply chain finance. The precision of their decision has so many degrees of importance for their business. Then Assessment of the credit worthiness is so important and they look for different quantitative models in order to reach to the credit scoring of their customers. Some of the reasons Refah-Kargaran Bank in Iran has been faced with the high amount of nonperforming loans are as following points. First one is Granting credit to the related parties of the board members and stockholders. On the other hand, the data of the improvements in the credit behavior is updated with a delay. According to the risk knowledge, all related behavior of depositors must be analyzed in credit scoring and Data of All financial institutions and banks and centers which give credit facilities should be analyzed. On the other side, some decision making dashboards do not exist and also the amount of risk must be assessed and assigned. Finally, the amount of collateral should be assigned according to the level of risk and of course, there is no information regarding the ability to pay the installment by customers [1].

This study offers a new approach for feature screening in the clustering of massive datasets, in which the number of features and the number of observations can be numerous. Benefitting a fusion penalization based on convex clustering criteria, we suggest a highly scalable screening procedure that efficiently discards no informative features by first computing a clustering score corresponding to the clustering tree constructed for each feature, and then thresholding the resulting values. We present theoretical support for our methodology by establishing uniform non-asymptotic bounds on the clustering scores of the "noise" features. These bounds imply perfect screening of non-informative features with high probability and are derived through careful analysis of the empirical processes corresponding to the clustering trees that are constructed for each of the features by the associated clustering procedure [2].

Clustering is the learning where the items are grouped on the basis of some inherent similarity. There are different methods for clustering the objects such as hierarchical, partitional, grid, density based and model based [3].

During the last two decades, the financial crisis in the banking system has damaged many banks and financial institutions around the world, causing some of them to go bankrupt. As a result of such crises, it became critical for observers to identify the sources of crises in order to reduce their intensity and impact. In 2015, the total Nonperforming loans reached 200 trillion Rials (about US $5.5 billion). According to the latest report by the Central

International Journal of Supply Chain Management

IJSCM, ISSN: 2050-7399 (Online), 2051-3771 (Print)

Bank of Iran, these loans make up 15 percent of the country's total liquidity. Total deferred loans in the banking system have been increasing since 2011, indicating that the macroeconomic conditions of the country can have a significant impact on bank credit portfolio and loan quality. Financial crisis can encourage depositors to withdraw their savings. If depositors sense the uncertainty and volatility in the banking system and find better ways of investing their money, they begin to withdraw their savings from banks. In addition, banks lend a major portion of customers' savings; if these loans are not repaid when they are due, banks will experience a sudden drop in their resources and, in worse cases, can face bankruptcy. In general, financial crisis refers to a shock or a sudden change in most or all financial indicators, including short-term interest rates, asset value, changes in management behavior and performance, bankruptcy, and the collapse of financial institutions [4].

This study delves into the practice of credit scoring and introduces the use of the clustered support vector machine (CSVM) for credit scorecard development. This recently designed algorithm addresses some of the limitations noted in the literature that is associated with traditional nonlinear support vector machine (SVM) based on methods for classification [5].

2. Significance of the problem

Increase in loan default reduces the banking system's resources. Of course, given the various incentives for lending, a portion of deferred loans can be reabsorbed into the banking system as deposits, and banks pay their interest by the rate of ordinary deposits. On the other hand, reduction in resources reduces the lending ability, and thus the profitability, of banks. This, in turn, can significantly increase the interest rate on deposits [6].

Moreover, a bank loan default can have certain effects on the economy: Expansion of the volume and scope of the informal economy; Flow of deferred payments to other financial markets and real assets with speculation motives, including capital, housing, automobile, and gold markets; Change in the volume of liquidity and increased volume of money supply; Increased debt of the banking system to the Central Bank; Change in the composition and volume of the monetary base and injection of high-powered money into the economy; Increased interest rate on loans and a reduction in investment; Increased interest rate on deposits and increase in savings [7].

The banking network connects depositors. In Iran, banks and the monetary market play the major role in financing economic activities. Therefore, the increased outstanding debt of banks directly affects their performance and reduces their productivity. It also seriously damages the national economy and exacerbates the financing crisis of economic entities [8].

Increase in volume of deferred loans reduces the resources of the banking system, while reducing their lending ability and affect their money creation, thus reducing the volume of money supply [9].

As the resources of the banking system drops, banks are no longer able to respond to demands for loans at the current interest rate. Banks have three solutions to this problem:

1. Compensate for the lack of resources for lending by borrowing from the Central Bank;
2. Increase the interest rate on loans and the cost of using money as well as the risk of lending [10];
3. Increase the interest rate on deposits in order to incentivize depositors and investors and attract more resources.

The first option increases the banking system's debt to the Central Bank as a component of the nation's monetary base, and increased high-powered money injected into the monetary market increases both the volume of liquidity and money supply (change in the composition and volume of the monetary base).

The second option increases interest rates on loans, which adds to the cost of using bank resources and turns cheap bank resources into expensive ones. From an investor's perspective, this is a warning sign, which discourages them from continuing productive economic activities and creating new opportunities for investment, production, and job creation.

The third option involves increasing interest rates on deposits, which forces banks to pay higher interest in order to attract resources. Investors and depositors are incentivized to save their money, thus increasing banks' resources. However, increased savings and lower investment due to higher lending rate create a gap between investment and savings [11].

3. Clustering

Classification of similar objects into several groups is an important human activity. In everyday life, this is part of the learning process. Due to fast technological development, the volume of data stored in databases is increasing at a rapid rate. Analyzing stored data and converting them into information that can be used by an organization requires powerful tools. In marketing, first customers are classified based on various indicators (variables). Then, the behavior of each segment is identified and plans are made to provide better and more specialized service to them [12].

A technique used in identifying the target market is market segmentation, which involves identifying homogenous subsets of the market by clustering customers based on a set of variables. Segmentation is the process by which consumers with similar needs and expectations are clustered in a distinct segment of the market. It is assumed that consumers making up a segment of the market are homogenous and distinct from other segments. Some researchers tend to identify the timing of different demands in their clustering. Clustering provides a more rational and accurate matching of products and marketing efforts to the needs of consumers. In other words, market clustering is the process of separating customers into several distinct segments, each of which has its own limited a unique set of demands.

Therefore, a market segment is an almost homogenous group of customers that respond similarly to a particular marketing strategy. There are two ways for clustering customers:

1. Classification, where the researcher selects a set of variables of interest and then clusters customers based on these variables.
2. Clustering, where the researcher selects a set of dependent variables and then divides customers into groups with high within-group similarity and average between-group similarity [13].

Effectiveness of clustering depends on whether or not the cluster (segment) is measurable, substantial, accessible, differentiable, and actionable. The company selects a number of clusters and treats them as smaller markets.

[14] Called the late nineteenth century the “segmentation phase”. He argued that a new area of micromarketing and over-clustering is being formed that is based on information technology [15]. This nascent area aims to create closer and more intimate relationships between producers and their target markets. Micromarketing is a form of targeted marketing by which organizations adapt their marketing (products, promotions, and efforts) to the very personalized needs of geographic, demographic, social, economic, psychographic, and benefit-based segments.

One of the advantages of clustering consumers to other techniques is that it is based on the assumption is that maintaining a current customer is cheaper than attracting a new customer; the reason for that is the long-term relationship with the customer, which creates value-added and loyalty [16].

We offer an ensemble classification approach based on supervised clustering for credit scoring. In this methodology, supervised clustering is employed to partition the data samples of each class into a number of clusters. Clusters from different classes are thus combined in pairs to form a number of training subsets. In each training subset, a specific base classifier is constructed. For a sample whose class label needs to be predicted, the outputs of these base classifiers are combined by weighted voting. The weight associated with a base classifier depends on the classification performance in the neighborhood of the sample. In the experimental study, two benchmark credit data sets are selected for performance assessment, and an industrial case study is conducted. The results show that compared to other ensemble classification methods, the proposed approach can generate base classifiers with higher diversity and local accuracy, and improve the accuracy of credit scoring [17].

4. Credit risk and credit scoring

According to the Basel Committee on Banking Supervision (BCBS), the main risks to which banks are exposed are: credit risk, country risk, transfer risk, market risk, interest rate risk, liquidity risk, operational risk, legal risk, and reputational risk [18]. Credit risk is one of the main risks that banks have to face. It is the risk of default on a debt that may arise from a depositor failing to make required payments due to unwillingness or financial disability. For any financial service provider like commercial banks, it is essential to separate good and bad customers. Thus, valid models are needed to predict potential default on loans so that stakeholders can take proper preventive and corrective measures.

5. Methodology

The present research is a descriptive survey. The population consists of the transactional and demographic data of the legal clients of Refah-Kargaran Bank. The sample consists of the data of 1000 randomly-selected clients. 80 percent of the selected sample has terminated their contract successfully with good credit status and remaining 20 percent has had a negative credit history. The purpose is to score the credit of individual depositors of this bank using clustering techniques for finance supply chain. The selected credit scoring methodology should have enough flexibility in order to be updated in varying and different economic conditions. For this purpose, the fuzzy clustering model is one of the best solutions according to its technical characteristics in continuously considering the different economic situations in its results [21], [22].

5.1. K-Means Clustering

K-means is the most widely used clustering techniques. It was introduced by James Mac Queen in 1967. In this technique, the number of clusters is determined at the onset. It is designed for clustering numerical data. The cluster has a centroid called the "mean". In K-means clustering, objects are randomly divided into clusters. Next, the distance from each object to the center is calculated. If the distance is greater than the cluster mean, the object is assigned to the closest centroid. This process is repeated until the error function is minimized and/or the centroid do not change (Momeni, 2011).

If D is a set of data with objects, and $\{C_1, C_2, \ldots, C_k\}$ denote separate clusters, then the error function (EF) is the sum of the distance from each object to the centroid of its cluster:

$$EF = \sum_{i=0}^{k} \sum_{x e c_i}^{.} d(x, \mathrm{M}(c_i)) \tag{1}$$

Where M denotes cluster centroid and is each object's distance from its own cluster mean.

5.2. FCM Clustering

In fuzzy C-mean (FCM) clustering, the number of clusters is determined at the onset. The objective function for this algorithm is defined as:

$$J = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^{m} d_{ik}^{2} = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^{m} \|x_k - v_i\|^2 \tag{2}$$

In this formula, is a real number greater than 1, and in most cases is given the value 2. Is the element and is the centroid of the cluster. Denotes the extent to which the element belongs to the cluster. Denotes the similarity (distance) of the element with (from) the cluster center, and any function that represents the similarity can be used. Matrix U can be defined on with rows and columns, and its elements take a value between 0 and 1. If all the elements of matrix U is 0 or 1, the algorithm will be similar to the classical c-mean algorithm. Although the elements of matrix U can take any value between 0 and 1, the sum of the elements in each row must be equal to 1:

$$\sum_{i=1}^{c} u_{ik} = 1, \forall k = 1, \ldots, n \tag{3}$$

This condition means that the sum of each sample's belonging to the cluster must be equal to 1. Using this condition and minimizing the objective function gives:

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^{m} x_k}{\sum_{k=1}^{n} u_{ik}^{m}} \tag{4}$$

Algorithm steps:
1. Initializing c,m, and U^0. Initial clusters are generated.
2. Cluster centers (v_i) are calculated.
3. Membership matrix is calculated based on the clusters from step 2.
4. If $||\mathrm{Ul} + 1 - \mathrm{Ul}||\ £\ e$, the algorithm is terminated; otherwise, it is repeated from step 2.

5.3. Subtractive Clustering

The mountain clustering approach is relatively plain and efficacious. Anyhow, its ability to compute highly extends with the dimension of the patterns due to the method capability to evaluate the mountain function on the whole grid points. For instance, a clustering pattern with four variables in which each dimension owns a resolution of 10 grid lines will result in 10 grid points which ought to be evaluated. Subtractive clustering is an alternative method that can be applied to diminish the above mentioned issue. In subtractive clustering, data points (not grid points) are regarded as the candidates for cluster centers. Through applying such method, the computation is easily relative to the number of data points and independent of the dimension problem as already [19].

Take a collection of n data points into consideration $\{x_1, \ldots, x_n\}$ in an M-dimensional space. The data points are supposed to have been normalized within a hypercube.

$$D_i = \sum_{j=1}^{n} \exp\left(-\frac{\|x_i - x_j\|^2}{\left({r_a}/{2}\right)^2}\right) \tag{5}$$

Because each data point is a candidate for cluster centers, a density measure at data point xi is defined as where r_a is a positive constant. Thus, a data point will have a high density value if it has many data points around. The radius r_a defines a neighborhood; data points outside this radius have effects only a bit on the density measure. After the density measure of each data points is calculated, the data point with the highest density measure is chosen as the first cluster center. Let x_{c1} be the point selected and D_{c1} its density measure.

$$D_i = D_i - D_{c1} \exp\left(-\frac{\|x_i - x_{c1}\|^2}{\left(r_b/2\right)^2}\right) \quad (6)$$

The density measure for each data point xi is revised by the formula where r_b is a positive constant. Therefore, data points near the first cluster center x_{c1} will have a highly decreased density measure, by which making the points unlikely to be selected as the next cluster center. The constant r_b defines a vicinity that has measurable reductions in density measure. The constant r_b is normally larger than r_a to stop closely spaced cluster centers; generally r_b is equal to $1.5r_a$. After the density measurements for each data point is revised, the next cluster center x_{c2} is selected and all of the density measures for data points are revised again. This process is repeated until enough number of cluster centers are created [20].

5.4. Group AHP

Analytical hierarchy process (AHP) is one of the most well-known techniques for multi-criteria decision-making. It was first introduced by [14] in the 1970s. AHP is a structured technique for organizing and analyzing complex decisions. A decision may involve several decision-makers, and all their views must be considered in pairwise comparisons. In these cases, the following formula is used:

$$a'_{ij} = \left\{\prod_{l=1}^{k} a_{ijl}\right\}^{1/k} ; l = 1,2,\dots,k \quad \text{Number of decision-makers}$$
$$i,j = 1,2,\dots,n \quad ; i \neq j$$

And if one decision-maker, given their expertise and responsibility, has a greater effect on the process, a weight is added to their opinion and the following formula is used:

$$a'_{ij} = \left\{\prod_{l=1}^{k} a_{ijl}^{w_l}\right\}^{1/\sum_l w_l} \quad (7)$$

It is better to include the opinion of different decision-makers in group calculations when their consistency ratio (CR) is less than 0.1. In group AHP, CR is calculated as follows:

$$CR = \frac{\lambda_{max} - n}{n} \quad (8)$$

6. Definition of variables

Selecting the right variables is essential to constructing a model. To cluster customers based on risk, the first and most important step is to identify risk factors. In this research, variables were selected in two stages. First, scientific articles in reputable journals were reviewed from 2010 to 2016. 28 variables were identified and extracted. These variables were classified into two groups based on personal characteristics and loan profile. Then, variables that may not have been obtainable in all cases were omitted, leaving 11 factors that are listed in Table 1.

Table 1-Factors obtained from literature review

Factors	
Personal Characteristics	Age
	Gender
	Location
	Marital status
	Education
	Job
Loan Profile	Creditworthiness
	Loan term
	Collateral type
	Collateral value
	Loan value

In the second stage, a questionnaire was designed and distributed among experts in the fields of finance and banking with enough experience and knowledge. In order to filter and organize the information contained in the research, the following procedure below was used:

1. Developing a conceptual model

2. Review and filtering of outlier and missing data
3. Preview of raw data for benchmarking
4. Using the automatic data categorization feature (Anomaly Index)

Here, data are automatically categorized using one of the features of the Clementine software (Figure 1).

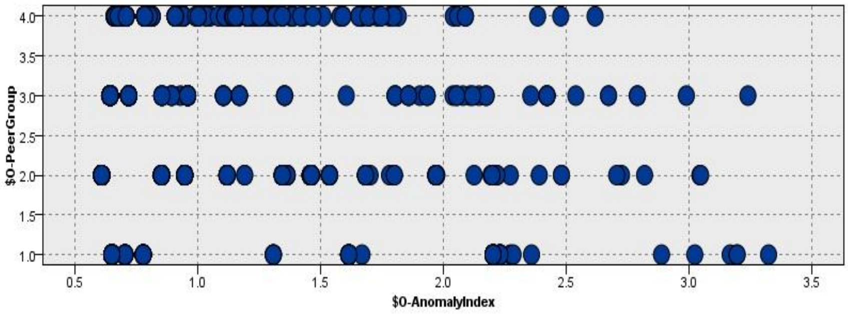


Figure 1. Representation of the automatic data clustering output for initial evaluation of inputs

5. Detecting and deleting outliers in the automatic data categorization index (Anomaly Index)

Here, data with values greater than 2.5 are removed from the model (Figure 2).

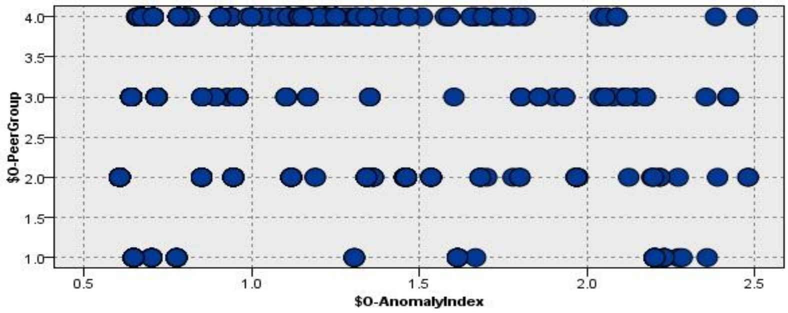
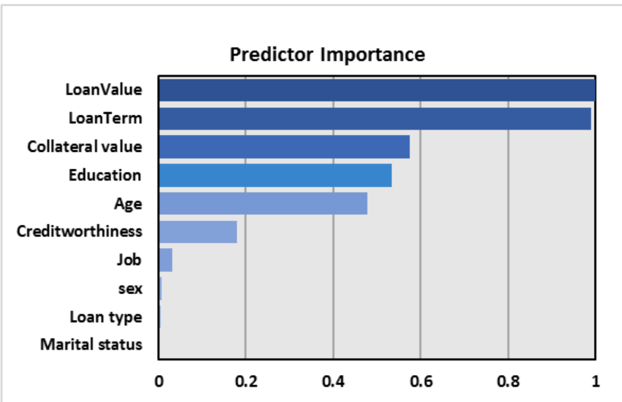


Figure 2. Representation of automatic clustering output after removing model noise

After examining a hundred different types of clustering using K-means, two Step, hierarchical, and Kohonen map and after reviewing the literature, K-means was found to be the most appropriate clustering technique for our purpose. It must be noted that various measures were taken to increase model accuracy, including the identification of important predictors and removing those that were less effective.

Next, the model was rerun to increase its accuracy. The resulting model is shown in the following figure. As can be seen, different clusters have a different credit status. Clusters can be divided into A, B, C, D, and E groups. Customers in cluster A, for instance, have all the characteristics of this cluster.



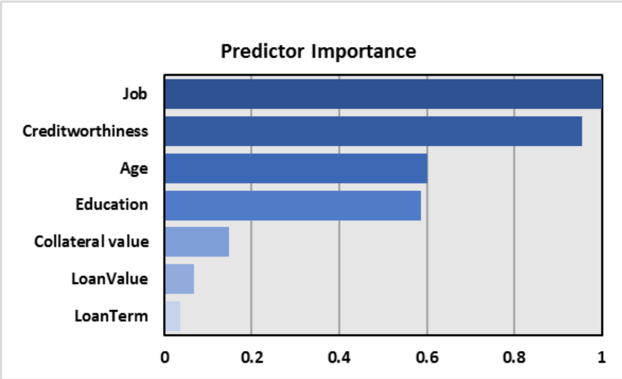


Figure 3. Identification and removal of less important features to increase model accuracy

7. Priority of credit scoring indicators

As shown in the conceptual model of the research, different factors affect customer evaluation. By prioritizing these factors, banks can credit depositors more effectively. The ranking and weight of each factor are shown in Table 2.

Table 2- Ranking and weight of factors with different initial starting points

Value	Count	Count	Count	Average	WEIGHT
Job	249	88	31	123	0.121
Loan Value	219	171	144	178	0.176
Sex	214	91	30	112	0.110
Marital Status	72	192	67	110	0.109
Loan Term	66	255	246	189	0.187
Loan Type	51	2	206	86	0.085
Collateral Value	49	171	243	154	0.153
Age	35	44	27	35	0.035
Education	32	13	3	16	0.016
Creditworthiness	13	6	3	7	0.007

It must be noted that group AHP was used to calculated the weight of these factors with inputs from eight banking experts and to convert coded data to quantitative data. The outputs of this procedure is as follows:

Table 3- Ranking and weight of factors for converting coded data

Credit worthiness	Loan Value	Collateral Value	Collateral Type	Loan Term	Job	Education	Age
0.38	0.15	0.23	0.47	0.13	0.38	0.43	0.28
0.32	0.40	0.32	0.32	0.36	0.19	0.44	0.31
0.17	0.28	0.21	0.22	0.51	0.16	0.14	0.14
0.13	0.17	0.23			0.09		0.26
					0.18		

Table 4- Ranking and weight of factors

Factor	Weight	Rank
Age	0.065	8
Education	0.140	3
Job	0.138	4
Loan Term	0.087	7
Collateral Type	0.104	6
Collateral Value	0.131	5
Loan Value	0.154	2
Creditworthiness	0.182	1

After determining the weight of factors and converting codes to numbers, clusters were created using K-means, FCM, and sub-clustering techniques in Clementine 18.0, MATLAB 2016, and Excel. About 35 models with different parameters were compared. Efficiency, predictive power, and separability of the clusters were evaluated using silhouette and Lift features.

Table 5- *Evaluation of clustering models using K-means and Sub clustering techniques*

		S* =	**0.192**	**0.298**	**0.372**
SUB Cluster			**LIFT 3**	**LIFT 4**	**LIFT 5**
	C1		0.195	0.200	0.184
	C2		0.169	0.162	0.153
	C3		0.221	0.228	0.221
	C4			0.164	0.164
	C5				0.252
		S* =	0.342	0.382	0.429
K-Means			**LIFT 3**	**LIFT 4**	**LIFT 5**
	C1		0.220	0.220	0.263
	C2		0.191	0.173	0.175
	C3		0.192	0.265	0.143
	C4			0.186	0.184
	C5				0.231

The below-mentioned figure illustrates the results of 100 times of simulations with different initial starting points for the customers and indicates that according to the LIFT index, five cluster model with FCM technique is the best one [20]. Evaluations show that K-means with five clusters is the most appropriate model. Moreover, the FCM technique was the most effective model in terms of separability. After removing nonessential variables, the model was rerun and the outputs were presented for each customer cluster. The results showed that creditworthiness, loan value, education, job, collateral value, collateral type, loan term, and age were respectively the most important variables.

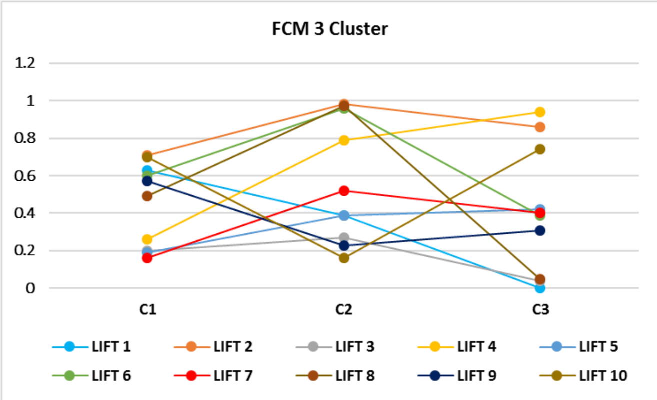


Figure 4. Evaluation of clustering model using (3 cluster), FCM techniques

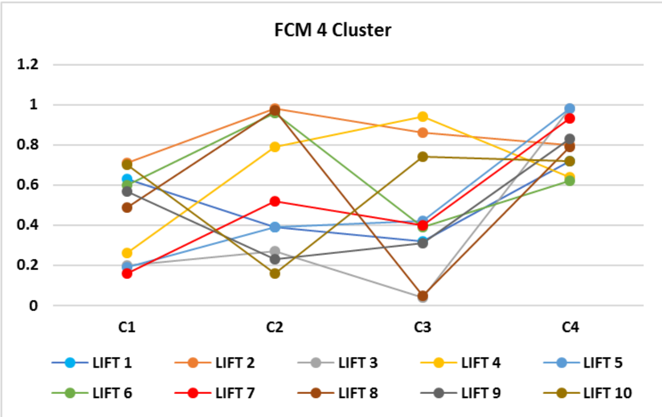


Figure 5. Evaluation of clustering model using (4 cluster), FCM techniques

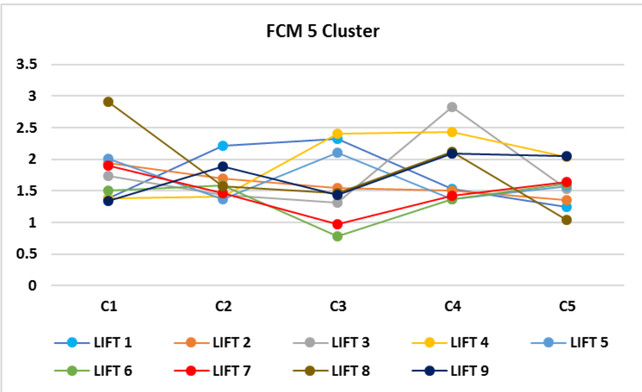


Figure 6. Evaluation of clustering model using (5 cluster), FCM techniques

8. Conclusion and recommendations

With new bank customers, it is necessary to collect their basic information based on the conceptual model of the present research and determine to which cluster they belong. Accordingly, a credit score is assigned to the customer which helps banks predict their creditworthiness and their ability to repay loans. In this research, the credit scoring model is based on coded data from Refah-Kargaran Bank of Iran, which can be used by tellers to make appropriate lending decisions based on credit scores and risk-taking strategies of the bank.

It is recommended to use the proposed model with a greater volume of data from the banking system to increase its accuracy and reliability and enhance its learning. Obviously, this can lead to more accurate lending decisions by banks.

References

[1] Amiri, A. *"Challenges facing the banking system and its impacts, Proceedings of the 1st National Conference on the Role and Function of the Banking System in Realizing Economic"*, Social, and Cultural Development Goals, Tehran, 2004.

[2] Atashkari, K., Nariman-Zadeh, N., Gölcü, M., Khalkhali, A., and Jamali, A. *"Modelling and multi-objective optimization of a variable valve-timing spark-ignition engine using polynomial neural networks and evolutionary algorithms"*, Energy Conversion and Management, Vol 48, 1029-1041, 2007.

[3] Bafande, A., and Rahimi, R. *"An expert fuzzy system for scoring bank customer credits"*, Iranian Journal of Commerce, 73, 1-27, 2014.

[4] Basu, S., Bilenko, M., Banerjee, A., and Mooney, R.J. *"Probabilistic Semi-Supervised Clustering with Constraints"*, Journal of Machine Learning Research, 71–98, 2006, http://www.cs.utexas.edu/~ml/ papers/semi-bkchapter-06.pdf.

[5] Central Bank of Iran. *A Summary of Economic Developments in 1990 Iran*. Tehran: Central Bank of Iran, Office of Economic Studies, 2013.

[6] Desai, V.S., Crook, J.N., and Overstreet, G.A. *"A comparison of neural network and linear scoring models in credit union environment"*, European Journal of Operational Research, 95, 24-37, 1996, http://www.sciencedirect.com/journal/european-journal-of-operational-research/vol/95/issue/1.

[7] Hanafizade, P., and Paidar, N.R. *"A model for risk classification of customer groups in car insurance industry using data mining techniques*", Iranian Journal of Insurance, Vol 26, 55-81, 2011.

[8] Harris, T. *"Credit scoring using the clustered support vector machine"*, Expert Systems with Applications, Vol 42, No. 2, 741–750, 2015, http://dx.doi.org/10.1016/j.eswa.2014.08.029.

[9] Hashemi, M., and Alimadad, M. *"Causes of deferred and overdue loans in the Housing Bank of Iran (1996-1997)*", Tehran: High Institute of Banking, 1998.

[10] Hosseinzade, L. *"Classifying target customers in the insurance industry using data mining*", Masters Dissertation, Tarbiat Modares University, 2007.

[11] Keshavarz, G.H., and Ayati, H. *"Comparison of the functioning of logit model and classification and regression tree models in credit scoring of bank customers: A Case of the Housing Bank of Iran"* Iranian Journal of Economic Research, Vol 7, 71-97, 2007.

[12] Khodavardi, O. "*Intelligent credit scoring: The case of an export credit agency*", Masters Dissertation, University of Tehran, 2009.

[13] Lacher, R.C., Coats, P.K., Sharma, S., and Fant, L.F. *"A neural network for classifying the financial health of a firm"*, European Journal of Operational Research, Vol 85, 53-65, 1995, https://doi.org/10.1016/ 0377-2217(93) E0274-2.

[14] Modares, A., and Zekavat, M. "*Credit risk models for bank customers"*, Iranian Journal of Auditing, Vol 19, 54-58, 2007.

[15] Moertini, V.S. *"Introduction to five data clustering algorithms"*, Integral, Vol 7, No. 2, 87-96, 2002.

[16] Momeni, M. "*Data Clustering (Cluster Analysis)*", Tehran: Moallef, 2011.

[17] Mooi, E. and Sarstedt, M. *"A Concise Guide to Market Research*", Berlin: Springer- Verlag Berlin Heidelberg, 2011.

[18] Olivan, A.D., Pagan, J.A., Sanz, R., and Sierra, B. *"Data-driven Prognostics Using a Combination of Constrained, K-means Clustering, Fuzzy Modeling and LOF-Based Score"*, Neurocomputing, Vol 241, pp. 97-107, 2017. https://doi.org/10.1016/j.neucom.2017.02.024.

[19] Safdari, N. *"Enhancing the banking system by covering the risk caused by deferred debts"*, The Latest in Economics (Iranian Journal), 122, 66-70, 2008.

[20] Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O.P., Tiwari, A., Joo Er, M., Ding, W., and Lin, CH. A "Review of Clustering Techniques and Developments*"*, Neurocomputing, Vol 267, 664-681, 2017, https://doi.org/10.1016/j.neucom.2017.06.053.

[21] Geunes, J., and Pardalos, P.M. *"Network optimization in supply chain management and financial engineering: An annotated bibliography"*, Networks: An International Journal, Vol 42, No. 2, pp. 66-84, 2003.

[22] More, D., and Basu, P. Challenges of supply chain finance: A detailed study and a hierarchical model based on the experiences of an Indian firm, Business Process Management Journal, Vol 19, No. 4, pp. 624-647, 2013.